REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY) 28-08-2012	2. REPORT TYPE Abstract	3. DATES COVERED (From - To) -

4. TITLE AND SUBTITLE
Terrorist Activity Pattern Detection in Afghanistan:A Knowledge Discovery and Data Mining Approach for Counter-Terrorism

5a. CONTRACT NUMBER
W911NF-11-1-0174

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
206022

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS
Jose Pou, Dr. Jeff Duffany (Advisor), Dr. Alfredo Cruz, (Mentor)

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Polytechnic University of Puerto Rico
377 Ponce De Leon
Hato Rey
San Juan, PR 00918 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
58924-CS-REP.2

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Data mining is primarily used by businesses. Today companies with strong consumer focus depend on data mining to determine relationships among internal and external data factors that are being registered and stored digitally. It enables them to drill down data into summary information to view detail transactional data and reveal trends that could be beneficial for an organization's business decision making. In this paper we incorporate data mining techniques using open source applications to model, evaluate and identify patterns of terrorism activity in

15. SUBJECT TERMS
Data Mining, NCTC, WITS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Alfredo Cruz
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		19b. TELEPHONE NUMBER 787-622-8000

Report Title

Terrorist Activity Pattern Detection in Afghanistan:A Knowledge Discovery and Data Mining Approach for Counter-Terrorism

ABSTRACT

Data mining is primarily used by businesses. Today companies with strong consumer focus depend on data mining to determine relationships among internal and external data factors that are being registered and stored digitally. It enables them to drill down data into summary information to view detail transactional data and reveal trends that could be beneficial for an organization’s business decision making. In this paper we incorporate data mining techniques using open source applications to model, evaluate and identify patterns of terrorism activity in Afghanistan for counter-terrorism and to strengthen homeland security. We apply data mining techniques to real terrorism incidents data from the Worldwide Incidents Tracking System (WITS) of the National Counterterrorism Center (NCTC). The results of the study will help in the discovery of terrorist group tendencies based on specific incident factors, but will also help evaluate the war on terror in Afghanistan up to date. With these results we also look to uncover valuable information regarding terrorist hot spots to determine geographical mobilization of security forces resources in the region.

Terrorist Activity Evaluation and Pattern Detection (TAE&PD) in Afghanistan: A Knowledge Discovery and Data Mining (KDDM) Approach for Counter-Terrorism

Data mining (DM) is primarily used by businesses to discover customer tendencies to guarantee future profit opportunities. In the TAE&PD project we intend to incorporate a KDDM methodology using open source applications to gather, preprocess, model, evaluate and identify patterns of terrorism activity that may prove useful to counter-terrorism and strengthen homeland security in Afghanistan. We will experiment using real terrorism incidents data from the Worldwide Incidents Tracking System (WITS) of the National Counterterrorism Center (NCTC). The project seeks to discover terrorism trends based on specific incident factors, help in the evaluation of war in Afghanistan and demonstrate a KDDM approach that could be applied (proof of concept) to national security. Project results may uncover valuable information regarding terrorist hot spots to determine geographical mobilization of security forces resources in the region.

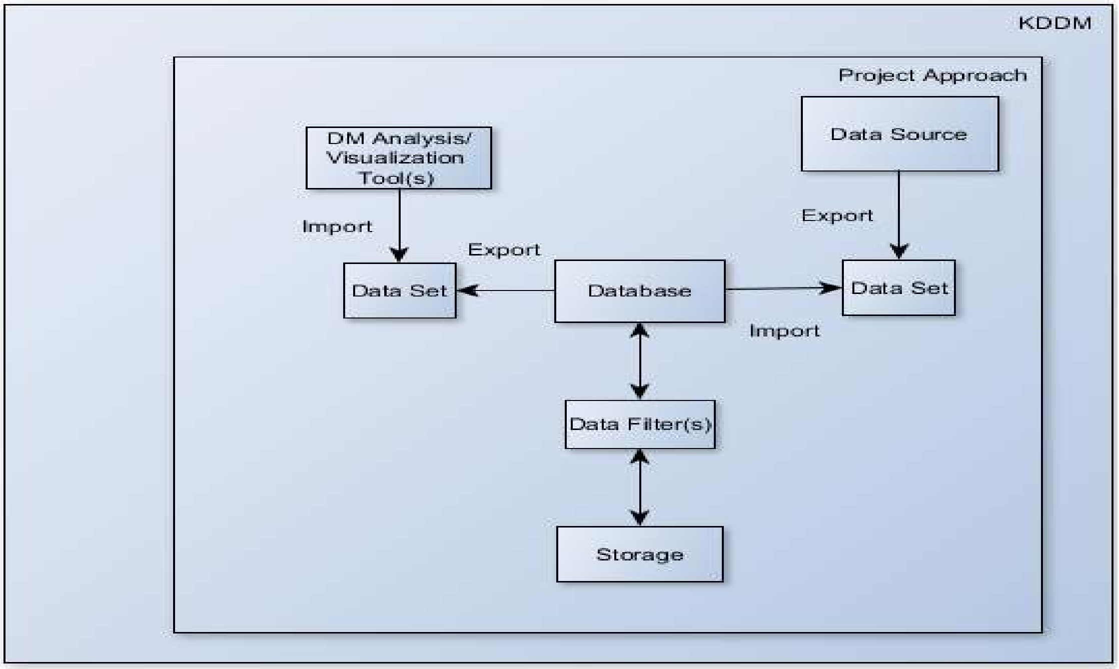

Figure 1: TAE&PD component's and data flow interaction diagram.

The following is a brief status report regarding the TAE&PD project current initiatives before implementation. It is crucial beforehand to prepare data accordingly as some attributes may not be suited for DM algorithms being considered. In this project we look to demonstrate the use of tools and techniques that are applied in KDDM related fields:

- Databases
- Statistics
- Machine Learning
- Visualization

The following table presents the current scope of project pertaining to KDDM field tools and techniques evaluation status that are still being researched before development of testing environment:

KDDM Field	**Tool(s)**	**Technique(s)**	**Status**
Databases	Microsoft SQL Server 2008 R2	Data Preprocessing • Cleaning • Integration • Transformation • Reduction	On going • New incident attributes (Integration) are being considered from other sources that may add more value to study. • Dataset contains incidents from 2004 to 2011. WITS site is offline for some time, no 2012 data can be collected for the moment.
Machine Learning	Microsoft SQL Server 2008 R2 Data Mining Add-ins.	• Clustering Analysis - Afghanistan incident type volume evaluation by country regions.	In progress • Researching - Acquired book Data Mining with Microsoft SQL Server 2008
Statistics	Microsoft SQL Server 2008 R2 Data Mining Add-ins.	• Time Series Analysis - Prediction of 2012 future incidents regarding deaths, injuries and kidnappings - Explore algorithms provided by tool.	Not started • Researching - Acquired book Data Mining with Microsoft SQL Server 2008
Machine Learning	R language/ Rattle	• Association Rules Analysis - Create rules associating incident month, week and province to type of attack. - Explore algorithms provided by tool.	In progress • Researching - Acquired book Data Mining with Rattle and R ()

Visualization	R language/ Rattle	• Graph representation of trends of incident data via the R language. - Security Forces (Police and Military) incident victim status by province or other factors.	Not started • Researching - R packages ~ RODBC ~ ggplot2 ~ arules ~ RStat

Table 1: TAE&PD project's current scope based on KDDM associated fields. S

The table above represents an initial blue print of the project in order to establish a defined scope.

After evaluating software used for data analysis the likes of R, Weka and Rapidminer it was concluded that the R language is the best fit for this project. R has the capability to be customized to the needs of any user and as far as DM use, it can be used by people with or without a programming background. R also has the advantage of a vast community of resources in comparison to other DM suites.

In the months of mid May, June and beginning of July of 2012 an effort has being made to acquire material to dive in R and its KDDM capabilities. R topics, to name a few, that have being studied include:

- Basic Operations
- Function(s)
- Introduction to Data Structures
  - Arrays
  - Lists
  - Data Frames
- Basic Charts and Graphs
- R Environment Creation
- Rattle and Data Mining
- Evaluation of Sampling Strategy
  - Training Dataset
  - Validation Dataset
  - Testing Dataset

In the coming month an effort will be made to start experimenting with Cluster Analysis method using SQL Server and Association Rules Analysis. Also experiment with the RODBC package to be able to interact directly with the TID database in SQL Server from R in order to present the use of the ggplot2 package for visualization.